

# Labour Force Survey: the impact of using an alternative imputation method on main labour market estimates

Labour Force Survey (LFS) person-weighted datasets use a roll-forward imputation method. During the coronavirus (COVID-19) pandemic an alternative imputation method that might better cope with the unprecedented labour market circumstances was explored. This article highlights the impact of using this alternative method on main labour market estimates over the periods between January to March 2018 and June to August 2022.

Contact:  
Nathan Compton  
labour.market@ons.gov.uk  
+44 1633 455400

Release date:  
24 October 2022

Next release:  
To be announced

## Table of contents

1. [Main points](#)
2. [Overview](#)
3. [Findings](#)
4. [Future developments](#)
5. [Nearest neighbour methodology](#)
6. [Related links](#)
7. [Cite this methodology](#)

# 1 . Main points

- The impact of the alternative imputation methodology on the story of headline labour market statuses (employed, unemployed or economically inactive) during the coronavirus (COVID-19) pandemic was minimal, with a maximum difference compared with published estimates of 0.4 percentage points, at or within sampling variability for all periods.
- The impact of the alternative imputation on headline labour market statuses in 2022 and before the coronavirus pandemic was even smaller, with the difference not exceeding 0.3 percentage points for any periods.
- The alternative imputation had a larger impact on hours worked estimates during the coronavirus pandemic, and this impact was not equally distributed between industries.
- The impact of the alternative imputation on hours worked has reduced in 2022, with the difference compared with published estimates of a similar size to that seen before the coronavirus pandemic.
- As the impact has now reduced, we will not continue to publish average hours estimates based on the alternative imputation method in our HOUR03: Average hours worked by industry dataset beyond the November 2022 release, and we will revert to publishing this dataset on a quarterly basis.

## 2 . Overview

Labour Force Survey (LFS) person-weighted datasets use a roll-forward imputation method. If we have a previous response, followed by a period of non-response, the previous response will be rolled forward to be used in the next period. This roll-forward is only allowed for one period.

The basis for this method is, although things change, responses in a previous quarter are a very good indicator of what that person would be doing in the next quarter. We only roll-forward for one quarter because by a second quarter, six months after the last recorded interview, circumstances become much more likely to have changed. If we fail to get a response for a second consecutive quarter, the respondent is dropped from the survey.

Another advantage with the roll-forward method is that it gives a fully cohesive record for the period. It has variables such as labour market status, industry, occupation, earnings, hours and so on, all being sensibly aligned with one another, because they all come from the same interview for that specific person.

### Issue

Because of the coronavirus (COVID-19) pandemic causing rapid changes in the labour market, what someone was doing in a previous quarter may no longer have been a good predictor of what they were doing in the current quarter. The primary areas of concern were labour market statuses, because of the rapid job losses, and changes in working hours, caused by the shutdown of certain industries and the introduction of the furlough scheme during the initial period of the coronavirus pandemic.

We decided to explore an alternative imputation method that might better cope during unprecedented labour market shocks, such as the ones brought about by the coronavirus pandemic. As this was a new, experimental method, it did not replace the roll-forward imputation method used to produce our published LFS estimates. Instead, we have used it to provide additional information during the coronavirus pandemic and to test its suitability as a potential replacement for the roll-forward method in the future.

The impact of using this alternative method on main labour market estimates for periods between January to March 2019 and May to July 2021 was highlighted in the [Labour Force Survey: alternative imputation during the coronavirus pandemic methodology article, published 2 November 2021](#). In this methodology article we further explore the impact of the alternative imputation method, by expanding the time series to cover the periods between January to March 2018 and June to August 2022.

The estimates in this new analysis have been updated to account for the latest LFS weights. Further details can be found in the [Impact of reweighting on Labour Force Survey key indicators: 2022 article](#).

## Method

Since June 2020, we have been publishing an alternative method involving donor imputation for average hours worked by industry in the [HOUR03: Average hours worked by industry dataset](#).

Instead of rolling responses forward from the previous quarter, missing values because of non-response are imputed from another respondent, referred to as the "donor". Where data would have been rolled forward, a suitable donor who returned data in the current period is searched for and their responses used to impute for the missing values. Missing respondents whose responses were to be imputed – referred to as "recipients" – are matched to donors using a "nearest neighbour" approach.

Potential suitable donors are identified using information on:

- age
- sex
- geography
- labour market status
- industry
- occupation
- hours
- ethnicity

Details of the individual variables that were used and how they feed into the identification of a donor can be found in [Section 5: Nearest neighbour methodology](#).

Once a suitable donor is identified we use their responses for suitable variables relating to hours worked and employment status, and some related variables (see [Section 5: Nearest neighbour methodology](#)). All other variables continue to be rolled forward.

We limited the method to these main variables because imputing values for all variables would involve much greater complexity. For example, some variables, such as someone's educational status, would be more appropriate to be rolled forward than taken from a donor. Whereas other variables, such as industry, would be more appropriate to roll forward if the labour market status has not changed, but would be more difficult if someone were imputed to change labour market status from unemployed to employed.

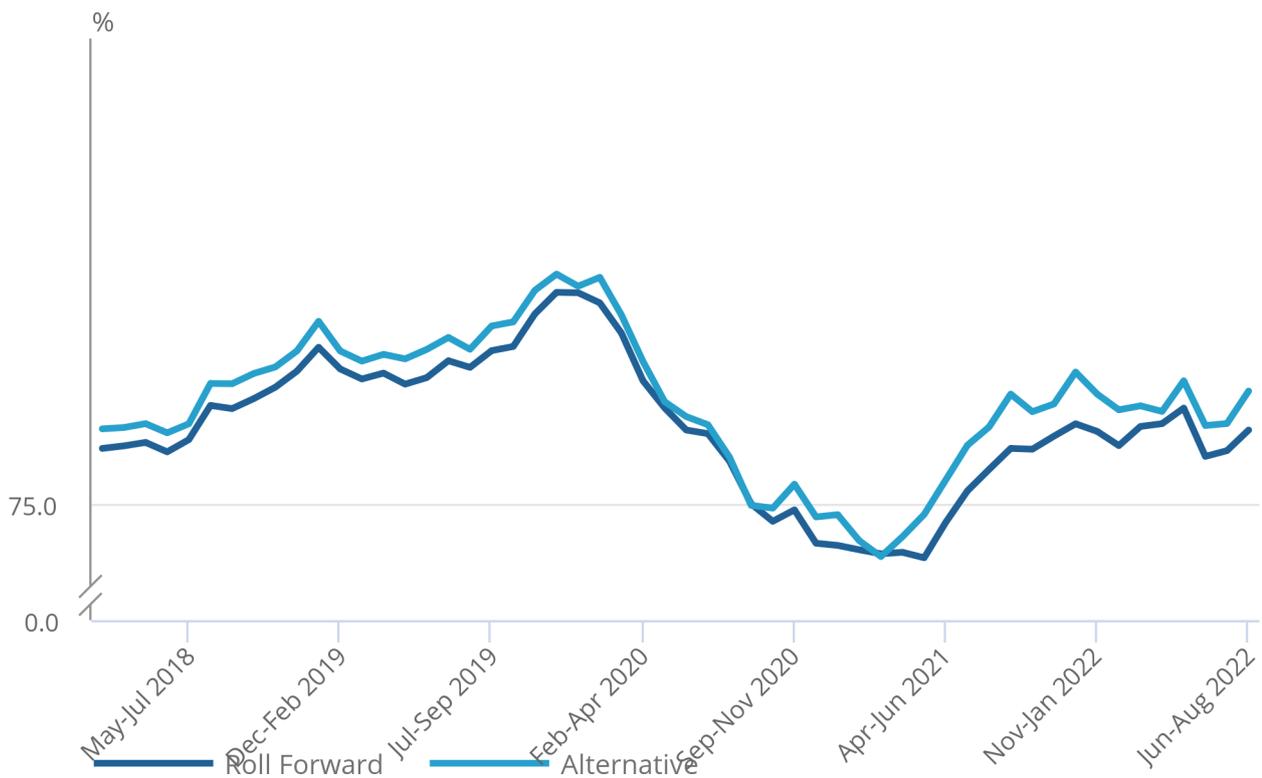
### 3 . Findings

**Figure 1: Employment rates have been minimally affected by the alternative imputation method, although the impact was larger during 2021**

UK employment rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022

Figure 1: Employment rates have been minimally affected by the alternative imputation method, although the impact was larger during 2021

UK employment rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022



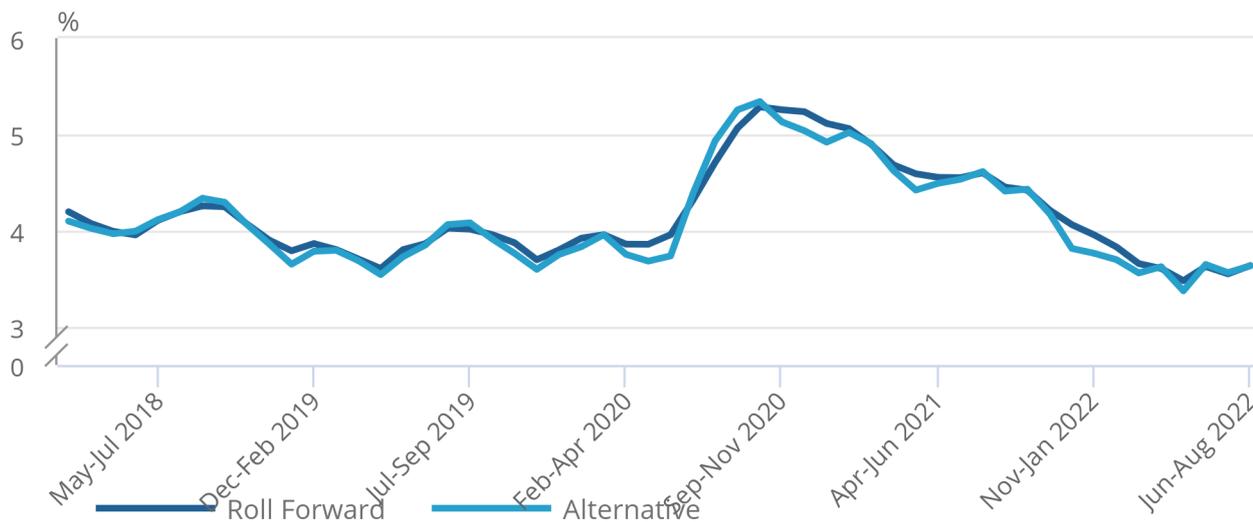
Source: Office for National Statistics – Labour Force Survey

**Figure 2: Unemployment rates have been minimally affected by the alternative imputation methodology**

UK unemployment rates, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022

Figure 2: Unemployment rates have been minimally affected by the alternative imputation methodology

UK unemployment rates, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022



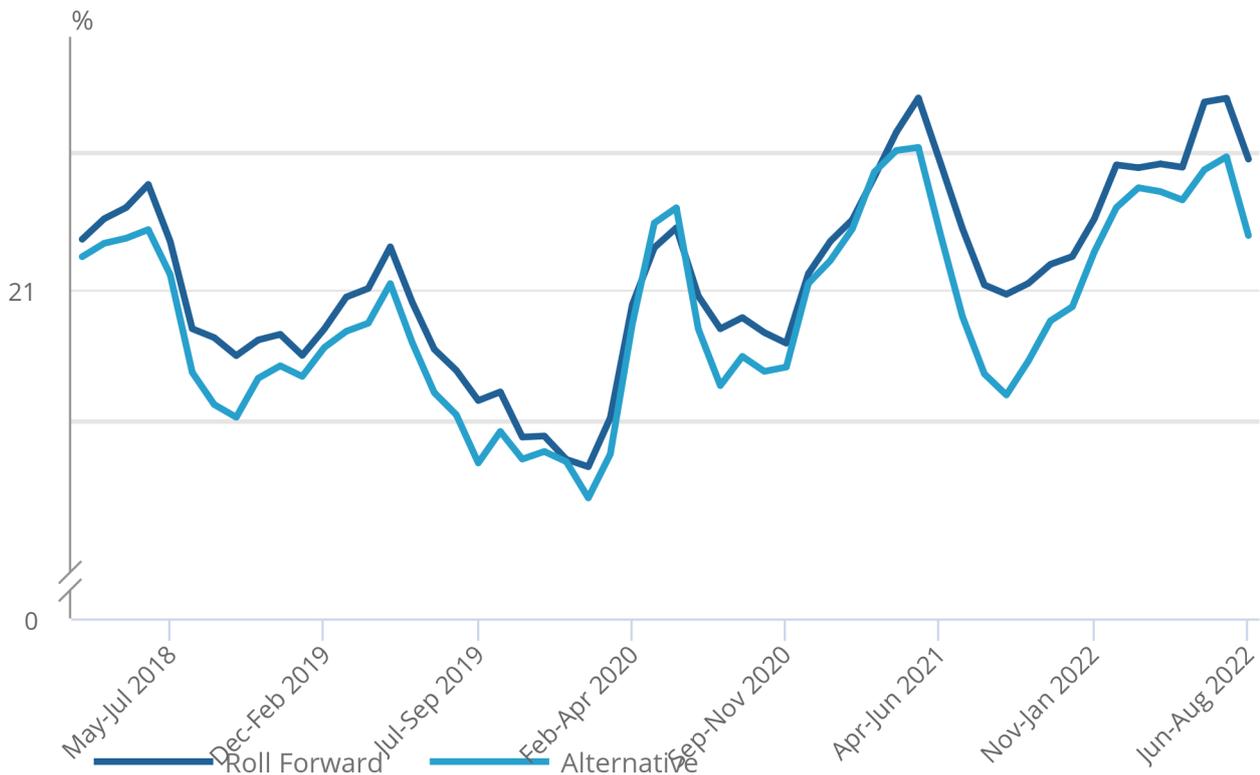
Source: Office for National Statistics – Labour Force Survey and Workforce Jobs

**Figure 3: Economic inactivity rates have been minimally affected by the alternative imputation method, although there was a larger impact during 2021**

UK economic inactivity rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022

Figure 3: Economic inactivity rates have been minimally affected by the alternative imputation method, although there was a larger impact during 2021

UK economic inactivity rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022



Source: Office for National Statistics – Labour Force Survey

The impact of the alternative imputation methodology on the headline labour market estimates between January to March 2018 and June to August 2022 was, overall, minimal.

Throughout the time-series, the employment rate estimates derived from the alternative imputation methodology were higher than those derived using the roll-forward method for all periods except two (July to September 2020 and January to March 2021, see Figure 1). Before the coronavirus (COVID-19) pandemic, the difference between the employment rate estimates derived from the two methods was small, with a maximum difference of 0.2 percentage points. The difference remained small during the initial stages of the coronavirus pandemic. It then became slightly larger in the first half of 2021, reaching a maximum of 0.4 percentage points in July to September 2021 and October to December 2021, before decreasing again in 2022.

The reason for the larger impact from using the alternative imputation method during 2021 was because nearest neighbour imputation responds more rapidly to extreme movements in the labour market, falling and rising earlier and more rapidly than roll-forward imputation, which introduces an element of inertia into the estimates. Despite this larger impact during 2021, the difference did not exceed sampling variability of the survey estimates (which was around plus or minus 0.4 to 0.5 percentage points in most periods; see our [A11: Labour Force Survey sampling variability dataset](#)).

For the unemployment rate, the inbuilt inertia from the roll-forward imputation results in a smoother series (Figure 2). The estimates derived using the alternative imputation are typically lower than those derived using the roll-forward method. This, however, is not consistent, with several periods throughout the time-series showing the opposite, particularly around mid-2020. Overall, the differences were not as affected by the coronavirus pandemic as they were for employment; they did not exceed 0.2 percentage points for the whole time series, which is at or within sampling variability for all periods (the sampling variability was typically around plus or minus 0.2 percentage points in most periods).

The story for economic inactivity counterbalances that of employment. The economic inactivity rate estimate derived using the alternative imputation method is consistently lower than that derived using the roll-forward method throughout the time series (with just a few exceptions), and the difference is larger during 2021 (Figure 3). Again, the maximum difference of 0.4 percentage points, seen in the periods June to August 2021 and July to September 2021, does not exceed sampling variability (which was plus or minus 0.4 percentage points in both of those periods).

**Figure 4: The alternative imputation method had a greater impact on total actual weekly hours worked than it did on the headline labour market statuses during the initial stages of the coronavirus pandemic**

UK total actual weekly hours worked, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2018 to June to August 2022

Figure 4: The alternative imputation method had a greater impact on total actual weekly hours worked than it did on the headline labour market statuses during the initial stages of the coronavirus pandemic



Source: Office for National Statistics – Labour Force Survey

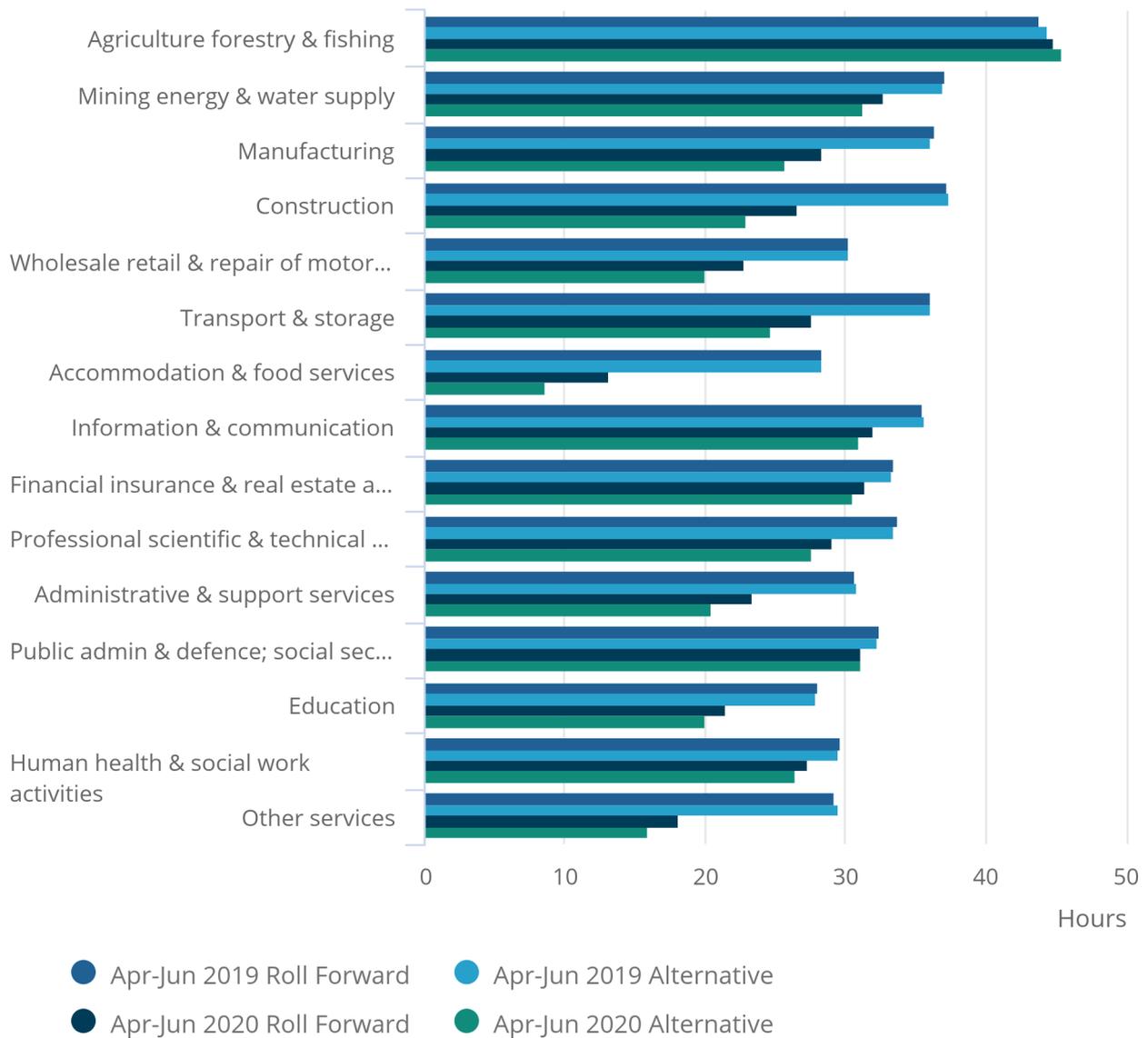
The alternative imputation method had a greater impact on total actual weekly hours worked estimates during the start of the coronavirus pandemic, where there was a sharper decrease in total hours than seen for total hours derived using the roll-forward method (Figure 4). Total hours were lowest in April to June 2020, and this was also when the difference in total hours derived from the two methods was largest: a difference of 63 million hours. The alternative methodology also suggested a faster recovery in total hours than the roll-forward method following the decrease, again, because of the inbuilt inertia from roll-forward imputation. The differences between total hours derived from the two methodologies decreased in the later stages of the coronavirus pandemic. In 2022, the differences have been of a similar magnitude to that seen before the coronavirus pandemic, typically less than 10 million hours each period.

**Figure 5: At the start of the coronavirus pandemic, average weekly hours worked in accommodation and food services was the most affected by the alternative imputation method**

UK average weekly hours by industry, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, April to June 2019 and April to June 2020

Figure 5: At the start of the coronavirus pandemic, average weekly hours worked in accommodation and food services was the most affected by the alternative imputation method

UK average weekly hours by industry, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, April to June 2019 and April to June 2020



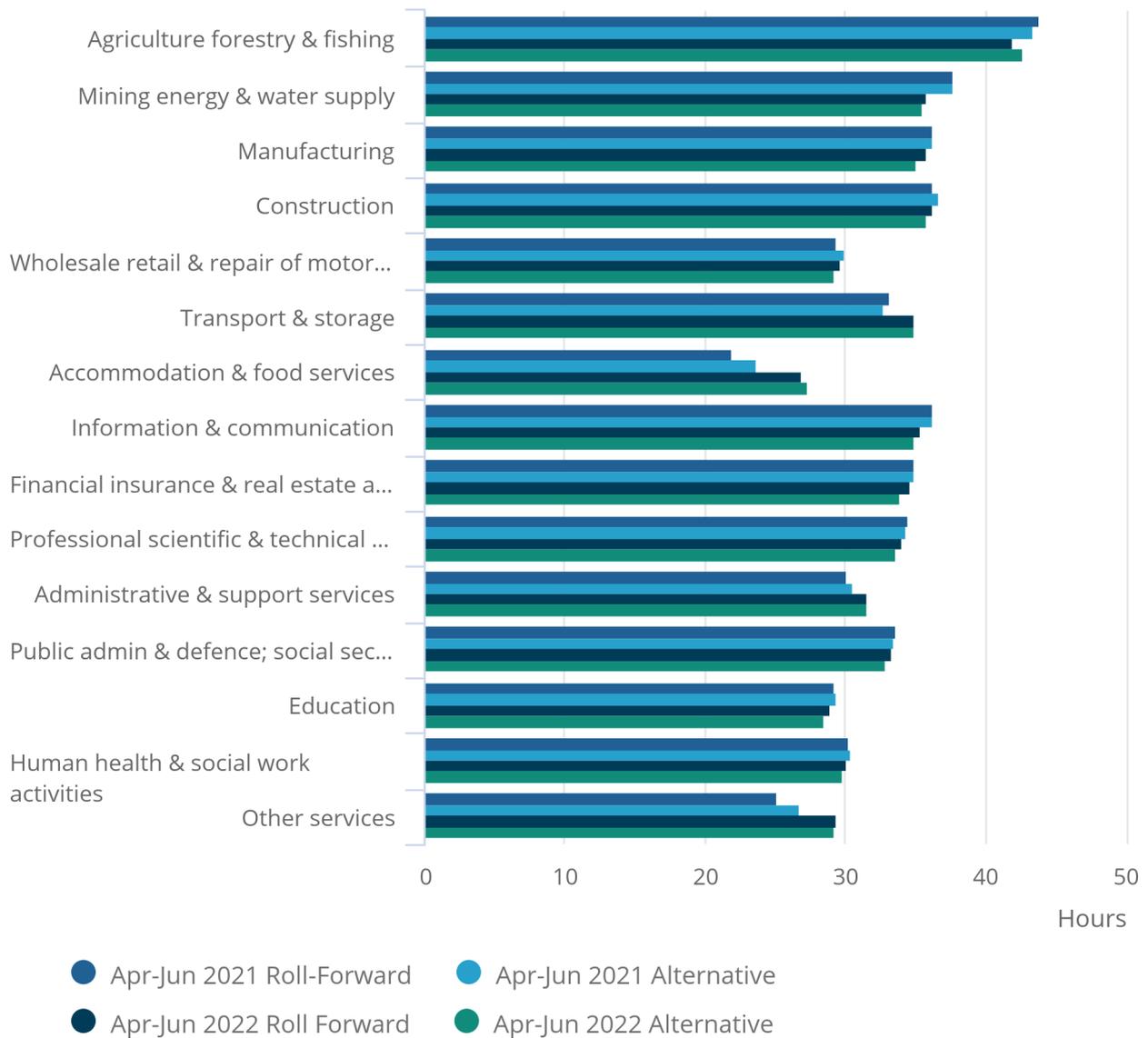
Source: Office for National Statistics – Labour Force Survey

**Figure 6: Differences in average weekly hours derived from the two methodologies decreased across almost all industries after the initial coronavirus pandemic period**

UK average weekly hours by industry, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, April to June 2021 and April to June 2022

Figure 6: Differences in average weekly hours derived from the two methodologies decreased across almost all industries after the initial coronavirus pandemic period

UK average weekly hours by industry, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, April to June 2021 and April to June 2022



Source: Office for National Statistics – Labour Force Survey

At the start of the coronavirus pandemic, the alternative imputation method for average actual weekly hours worked had an unequal impact on different industries. As shown in Figure 5, [industries most impacted by the coronavirus pandemic](#) were also those where the alternative imputation made the biggest difference (estimates are also published in [HOUR03: Average hours worked by industry dataset](#)). Average hours were at their lowest in April to June 2020, largely as a result of the furlough scheme. Many people in the most affected industries would have been on this scheme, in employment but working zero hours. In this period, accommodation and food services had the largest difference in average weekly hours, with average hours derived from the alternative imputation method being 33.8% lower than average hours derived from the roll-forward method. Construction saw the next largest difference, with the alternative method suggesting average hours were 13.5% lower than average hours derived from the roll-forward method. After the initial stages of the coronavirus pandemic, the differences in average hours derived by the two methods became smaller (Figure 6), with the differences in April to June 2022 not exceeding 2.2% for any of the industries. This is closer in magnitude to the differences seen before the coronavirus pandemic.

## 4 . Future developments

The alternative imputation has offered a different picture to help users understand the impact of the coronavirus pandemic on estimates of headline labour market statuses and hours worked. However, this analysis has demonstrated that, outside of the coronavirus pandemic period, using the alternative method has a relatively small effect on published estimates, with the differences in 2022 being of a similar size to those seen prior to the coronavirus pandemic. This, coupled with the complexities involved in incorporating this new methodology into existing systems, means that we do not intend to extend its use during the lifetime of the current Labour Force Survey. Therefore, the November 2022 release of our [HOUR03: Average hours worked by industry dataset](#) will be the last time we include estimates based on the alternative imputation method. In addition, HOUR03 will revert to being published on a quarterly basis, with the next release following the November release being in February 2023; this, and future releases, will only include estimates based on the roll-forward imputation method.

Further research into the alternative imputation methodology will continue, particularly regarding its potential application in the transformed Labour Force Survey, currently under development. For more information, see our [Labour market transformation – update on progress and plans: September 2022 article](#).

## 5 . Nearest neighbour methodology

The alternative imputation method involves donor imputation. Instead of rolling responses forward from the previous quarter, missing values because of non-response are imputed from another respondent who returned data in the current period, referred to as the "donor". Missing respondents whose responses were to be imputed – referred to as "recipients" – were matched to donors using a nearest neighbour approach. Potential suitable donors were first identified using the following variables:

- current AGE – variable derived separating respondents into groups of aged under 16 years, aged 16 to 64 years, aged 65 to 70 years and aged 70 years and over; this ensures a child aged under 16 years is not matched with a respondent aged 16 years and over
- current SEX – men and women
- current COUNTRY – resident country within the UK
- current GOVTOF – UK region of residence
- previous ILODEFR – economic activity
- previous INDS07M – industry section in main job (Standard Industrial Classification: SIC07)
- previous SC2010MMJ – major occupation group in main job (mapped Standard Occupational Classification: SOC10)
- previous STAT – employment status
- previous FTPT – full-time or part-time employment
- previous INECAC05 – detailed economic activity

Secondly, for each potential donor, "distance" measures were calculated for current age, ethnicity, and previous SUMHRS, as follows:

- current AGE:  $(\text{donor AGE} - \text{recipient AGE}) / 14$
- current ethnicity: if donor ETHUKEUL equals recipient ETHUKEUL (if the ethnicities match), then ethnicity distance equals 0, else ethnicity distance equals 1
- previous SUMHRS:  $\text{abs}(\text{donor previous SUMHRS} - \text{recipient previous SUMHRS}) / 97$

These distance measures using age, ethnicity and hours were then combined to provide an overall combined distance measure. The potential donor with the minimum distance measure, so the one that was most like the recipient, was selected as the donor. Once a suitable donor had been identified, their responses to the following variables were imputed to replace the recipients' missing data:

- BACTHR – total actual hours, main job, excluding overtime
- SUMHRS – total actual hours, main and second job, including overtime
- TOTHR – total actual hours, main and second job, including overtime
- TTACHR – total actual hours, main job, including overtime
- TTUSHR – total usual hours, main job, including overtime
- YLESS20 – reason for working fewer hours than usual in reference week
- WRKING – whether did paid work in reference week
- YTETJB – whether has paid job in addition to government training scheme
- TYPSC12 – type of work scheme
- FTPT – full-time or part-time employment
- ILODEFR – economic activity
- INECAC05 – detailed economic activity
- JBAWAY – whether temporarily away from paid work
- STAT – employment status

These variables were selected as the impact of this alternative imputation method on labour market status and hours worked was the priority at the time. Imputing values for all variables would involve greater complexity that, because of the urgency of the project, was not considered warranted at the time.

## 6 . Related links

### [Employment in the UK: October 2022](#)

Bulletin | Released 11 October 2022

Estimates of employment, unemployment and economic inactivity for the UK.

### [Labour Force Survey: alternative imputation during the coronavirus \(COVID-19\) pandemic](#)

Methodology | Released 2 November 2021

During normal operation, Labour Force Survey (LFS) person-weighted datasets use a roll-forward imputation. This article discusses an alternative imputation method that might better cope with the unprecedented labour market circumstances brought about in the pandemic and highlights the impact on main labour market estimates.

### [Impact of reweighting on Labour Force Survey key indicators: 2022](#)

Article | Released 23 May 2022

Indicative estimates of the LFS reweighting methodology on key indicators for the UK and countries of the UK, between January to March 2020 and October to December 2021.

# 7 . Cite this methodology

Office for National Statistics (ONS), released 24 October 2022, ONS website, methodology, [Labour Force Survey: the impact of using an alternative imputation method on main labour market estimates](#)